

Weekly Report

Period: 11/24/2013 – 12/01/2013

Projects

和小组成员讨论实现了 AVIS 0.0.1 版本。可参考 <http://10.76.0.182/AVIS/>，新增以下功能：

- ◆ 卫星云图动画
- ◆ 卫星云图数据可视化设置
- ◆ 联动交互

Research

【可视数据清洗】

自 2009 年以来，Jeffy Heer 等人以数据清洗为目标为研究目标，在 CHI 领域发表了一系列基于交互的可视数据清洗文章。目前，我大致看了一下这些文章，总结如下：

- [1] 解决的问题比较简单，基本值处理了数据重复、数值缺失等常见问题；
- [2] 无法处理大数据，只支持表格类数据的可视清洗；
- [3] 可视化不是这些文章的核心，这些文章主要是从交互角度进行的研究。

基于以上发现，我的基本思路：设计实现一个基于 Cloud Sourcing 的大数据清洗平台。首先，采用 D&R 的思路对数据进行切割。然后，采用可视化的方法清洗数据，生成一系列规则。最后，合并所有用户所产生的规则集合（这一步挑战比较大）。

可视数据清洗阶段，则分为两大步：一是语法类错误的处理（如：数据缺失，数据重复等）；二是语义类错误处理（如：特定环境下的异常检测）。期望能提出一些数据质量的度量标准，如正态性等。

目前，还没有完全理清楚这个研究领域所需要解决的核心问题。下周将继续研究这个领域内的文章，提出一些想法。

【信息论】

此外，本周我还主要在研究信息论的一些基础知识。下周将于气象局的领域专家询问研究问题的可行性。

Work to be done in next week

- 与张老师讨论 DTI 项目

- 提出一些可视数据清洗的基本想法

Reference: